

Molecular markers to assess genetic diversity of *Gentiana lutea* L. from the Ukrainian Carpathians

Maryana Z. Mosula¹, Igor O. Andreev^{2*}, Olena M. Bublyk², Vitaliy M. Mel'nyk², Iryna I. Konvalyuk², Nadiya M. Drobyk¹ and Viktor A. Kunakh²

¹Volodymyr Hnatiuk Ternopil National Pedagogical University, M. Kryvonis street 2, Ternopil 46027, Ukraine and ²Institute of Molecular Biology and Genetics NAS of Ukraine, Akad. Zabolotnogo street 150, Kyiv 03680, Ukraine

Received 21 May 2014; Accepted 15 November 2014 – First published online 12 December 2014

Abstract

The aim of the study was to develop the system of polymerase chain reaction (PCR)-based markers for the assessment of genetic diversity and population genetic studies of *Gentiana lutea* L. as well as to determine the utility of two indices of marker informativeness. The informativeness was determined for 40 PCR primers of different types (random amplified polymorphic DNA, inter simple sequence repeat, inter-retrotransposon amplified polymorphism, resistance gene analog polymorphism and conserved DNA-derived polymorphism markers) by evaluating discriminating power (D_L) and resolving power (R_p) in a sample of 30 plants from two populations. Analysis of correlation between the index value and the number of differentiated pairs of genotypes in the given sample revealed that D_L is more efficient than R_p ; therefore, we selected primers based on the D_L value. In total, 12 primers with the largest values of D_L were chosen. Analysis of genetic relationship among 86 plants from six populations showed that the number of bands produced by the three of selected primers was sufficient to give average bootstrap support across six key nodes in the dendrogram higher than 85%, while using six of the primers resulted in average bootstrap value exceeding 99%. Thus, a minimal set of three to six selected primers are sufficient for a quick assessment of genetic diversity of *G. lutea* populations, depending on the sample size and degree of differentiation between populations, while the rest of the primers with D_L values above 0.8 may be used for ecogenetic surveys. Preliminary results obtained with selected primers indicate the moderate level of genetic variation within the species and significant differentiation among individual populations.

Keywords: discriminating power; genetic diversity; *Gentiana lutea*; molecular markers; primer informativeness; resolving power

Introduction

Gentiana lutea L. (Gentianaceae) is a herbaceous perennial native to the mountains of Central, Eastern and Southern Europe. It grows in meadows and open

slopes, usually on calcareous soils, at altitudes ranging from 800 to 2500 m. Rhizomes of the plant contain numerous biologically active substances such as iridoids, alkaloids, xanthones, flavonoids, phenol carbonic acids and anthocyanins used as a remedy for fever, hysteria, high blood pressure and also to prevent muscle spasms (Jensen and Schripsema, 2002; Strashniuk *et al.*, 2006).

Due to over-exploitation of the natural resources, *G. lutea* is treated as an endangered species under protection in most

*Corresponding author. E-mail: i.o.andreev@imbg.org.ua

European countries, including Ukraine. The value of this medicinal plant and the potential demand for raw materials necessitate the assessment of its resources in Ukraine and finding possible ways for its preservation and restoration. To date, main populations of the species have been established and described in the Ukrainian Carpathians (Mayorova *et al.*, 2013). However, data on overall genetic resources and individual populations of *G. lutea* in Ukraine are still fragmentary (Mel'nyk *et al.*, 2004; Mosula *et al.*, 2013; Mosula *et al.*, 2014). Nowadays, polymerase chain reaction (PCR)-based methods of molecular–genetic analysis provide one of the most effective tools for studying genetic variation in plants and animals. However, study of a new object requires a search for optimal molecular–genetic markers and evaluation of their efficiency.

A number of indices have been proposed to assess the efficiency of PCR primers intended to be used in genetic analysis. Polymorphism information content (PIC) was first suggested as a parameter to evaluate informativeness of marker loci (Botstein *et al.*, 1980). Later, Powell *et al.* (1996) used the same index calculated from the frequencies of individual polymorphic bands (expected heterozygosity, H_e) along with multiplex ratio and marker index (MI) to evaluate the utility of different PCR-based marker systems. Prevost and Wilkinson (1999) have proposed resolving power (R_p) as a new measure of the ability of primers or techniques to distinguish between genotypes and demonstrated that MI failed to correlate with this ability. Finally, Tessier *et al.* (1999) suggested modification of PIC by calculating it from the frequencies of the different banding patterns generated by a primer and named it discriminating power (D_I). Several studies on various organisms have compared different indices of primer efficiency by the correlation with the primer's ability to distinguish between the pairs of individuals in a given sample set (Prevost and Wilkinson, 1999; Tessier *et al.*, 1999; Saini *et al.*, 2010). The results of these studies indicate that discriminating power and resolving power (R_p) have the highest correlation with the ability to distinguish individual genotypes. Furthermore, statistical approach proposed by Tessier *et al.* (1999) was used recently to develop computer program for marker choice (Caroli *et al.*, 2011). The objectives of our study were to determine the usefulness of these indices in the assessment of informativeness of PCR primers and to select the set of molecular markers for efficient investigation of genetic diversity of *G. lutea*.

Materials and methods

Collection of plant materials

Thirty accessions of *G. lutea* used in the study to evaluate indices of primer informativeness were collected from

two populations located on polonyna (mountain grassland) Krachuneska (Kr) and on the ridge slope between Troyaska and Tataruka Mountains (Tr) (Svydovets ridge, the Ukrainian Carpathians). These populations are growing in comparable ecological and geographical conditions as well as similar in population size and population density and mode of exploitation of natural environment (Mayorova *et al.*, 2013). To further evaluate the efficiency of selected primers on a larger sample, we used additional 56 samples of *G. lutea* from four other populations from the Chornohora ridge (Mayorova *et al.*, 2013) (Fig. 1).

DNA isolation and PCR amplification

DNA was isolated from fresh young leaves by the standard procedure (Rogers and Bendich, 1985). In total, ten random amplified polymorphic DNA (RAPD) primers and nine inter simple sequence repeat (ISSR) primers earlier used in gentian studies (Twardovska *et al.*, 2009; Konvalyuk *et al.*, 2011), nine conserved DNA-derived polymorphism (CDDP) primers and seven pairs of resistance gene analog polymorphism (RGAP) primers described in Collard and Mackill (2009) and Dong *et al.* (2009), and five inter-retrotransposon amplified polymorphism (IRAP) primers kindly provided by Dr R. M. Kalendar (MTT/BI Plant Genomics, Institute of Biotechnology, University of Helsinki) were used in PCR. Primer sequences are listed in Table S1 (available online).

Amplifications were performed in Tertsyuk MC2 thermocycler (Biotechnology, Russia). The 20 μ l PCR mixture contained 20–30 ng of genomic DNA, 0.2 mM each dNTP (Fermentas, Lithuania), 1.25 U *Taq* DNA polymerase (AmpliSens, Russia), 0.5 μ M of a primer, 1 \times PCR buffer with $(\text{NH}_4)_2\text{SO}_4$ and 2.5 mM MgCl_2 (Fermentas, Lithuania). Reaction mix was layered with a drop of mineral oil to avoid evaporation. As a negative control for amplification, a reaction mixture containing sterile water instead of DNA was used. Each reaction was performed at least twice. The amplification conditions were as follows: RAPD-PCR: 94°C, 2 min, five cycles (94°C, 30 s; 37°C, 30 s; 72°C, 1 min); 35 cycles (94°C, 20 s; 37°C, 20 s; 72°C, 40 s); 72°C, 2.5 min; ISSR-PCR: 95°C, 2 min, 35 cycles (94°C, 30 s; 53°C, 30 s; 72°C, 1.5 min), 72°C, 2.5 min; IRAP-PCR: 94°C, 2 min, 35 cycles (94°C, 30 s; 58°C, 30 s; 72°C, 1.5 min), 72°C, 2 min; CDDP-PCR: 95°C, 2 min, 35 cycles (94°C, 30 s; 53°C, 60 s; 72°C, 1.5 min), 72°C, 2.5 min; RGAP-PCR: 95°C, 2 min, 40 cycles (94°C, 30 s; 53°C, 45 s; 72°C, 1 min), 72°C, 2.5 min.

The PCR products were separated by gel electrophoresis in 1.3 % agarose gel in 1 \times sodium borate buffer

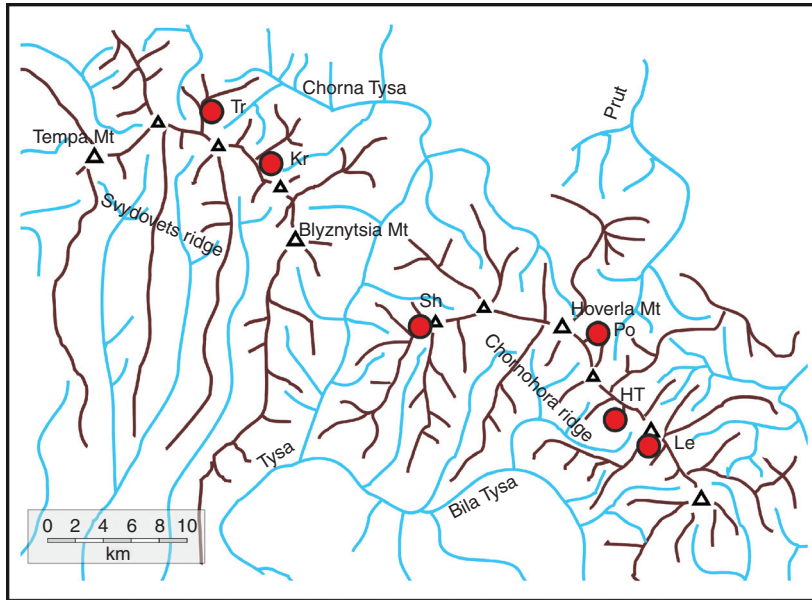


Fig. 1. Map of sampled populations of *Gentiana lutea* in the Ukrainian Carpathians (represented by grayed circles). Kr, polonyina Krachuneska; Tr, the ridge slope between Troyaska and Tataruka Mountains (Mts); Sh, Sheshul and Pavlyk Mts; Po, Pozhyzhevka Mt; HT, Hutyn Tomnatyk Mt; Le, polonyina Lemska.

(5 mM $\text{Na}_2\text{B}_4\text{O}_7$, pH 8.5), visualized by staining with ethidium bromide and photographed under ultraviolet light.

Statistical analysis

Amplification products were scored as 1 (present) or 0 (absent) for individual plant samples and binary matrix was generated. Only distinct, reproducible fragments were scored.

To evaluate the effectiveness of a primer, resolving power (R_p) (Prevost and Wilkinson, 1999) and discriminating power (D_L) (Tessier et al., 1999) were calculated as follows:

$$R_p = \sum lb,$$

where lb is the informativeness of a band that is determined based on the proportion of genotypes containing it (p): $lb = 1 - (2 \cdot |0.5 - p|)$.

$$D_L = 1 - \sum p_i^2,$$

where p_i is the frequency of the i th banding pattern generated by a primer. Also the number of non-differentiated pairs (ND) in a set of n genotypes was calculated for each primer from the frequency of generated banding patterns p_i (Tessier et al., 1999):

$$\text{ND} = (n \cdot (n - 1) / 2) \cdot \sum_{i=1}^n p_i \frac{n \cdot p_i - 1}{n - 1}.$$

For a given combination of k primers, under hypotheses of independence of the formation of banding

patterns for individual primers, this number (ND_k) is equal to:

$$\text{ND}_k = (n \cdot (n - 1) / 2) \cdot \prod_{j=1}^k \sum_{i=1}^n p_i \frac{n \cdot p_i - 1}{n - 1}.$$

Proportion of polymorphic bands (P_B), expected heterozygosity (H_e) and Shannon index (S) were calculated using GenAlEx 6.5 software (Peakall and Smouse, 2006, 2012). Jaccard's genetic distances (D_j) were calculated and unweighted pair group method with arithmetic mean (UPGMA) trees were constructed using FAMD 1.3 (Schluter and Harris, 2006). The robustness of the UPGMA dendrograms was assessed with bootstrap analysis running 1000 iterations using the WinBoot software (Yap and Nelson, 1996).

Results

A total of 40 primers of different types were chosen for the study based on the results of preliminary assessment of quality and quantity of amplification products generated with template DNA from *G. lutea* (see online supplementary Table S1). These primers were applied to analyse the genetic variation in a set of 30 plants. As *G. lutea* belongs to the species with fragmented range which are at high risk for significant differentiation of the populations, we tried to account for the spatial character of the species distribution and included in this set plants originated from two isolated populations in the Svydovets ridge of the Ukrainian Carpathians.

Based on the generated PCR banding patterns, a binary matrix was constructed, and indices of informativeness were calculated for individual primers. In particular, for each primer, we determined the total number of generated bands (N_t), the proportion of polymorphic bands (P_B), the resolving power (R_p), the discriminating power (D_L) and the number of non-differentiated pairs (ND) (see online supplementary Table S1). N_t ranged from 4 to 26 with a mean of 12.4, and P_B ranged from 0 to 100% with a mean of 74.4%. A comparison of N_t and P_B values calculated for various subsets of plants reveals the differences among the individual populations and the total sample of plants. D_L varied from 0 to 0.967 with a mean of 0.821, and R_p had values from 0 to 12.3 with a mean of 5.5. The number of non-differentiated pairs ranged from 0 to 465, and only one primer was able to differentiate between all analysed genotypes (see online supplementary Table S1).

There was only moderate correlation between R_p and D_L ($r = 0.611$, $P = 0.01$), that can be explained by the specificity of distribution of polymorphic bands, in particular by the fact that the products of the primers that have a large difference between R_p and D_L showed mainly between-population differences, while their within-population variation was quite low. Considering the moderate correlation between R_p and the number of non-differentiated pairs ($r = 0.610$, $P = 0.01$) compared with direct relationship between D_L and ND, we excluded it from further analysis and selected primers based on the D_L value.

In total, 12 primers with the largest values of D_L were chosen (Table 1). These included five ISSR, four RAPD,

two CDDP and one IRAP primers. Primer UBC#807 allows discrimination of all plants from two populations under study, while the others can be used for this purpose in a combination of two or three.

To analyse the influence of the number of used markers on variation of the measures of genetic diversity, we compared their values calculated for the sample of plants from two populations using all primers or various combinations of primers with the largest D_L as follows:

- (1) 40 primers – all primers used in the study;
- (2) 12- D_L – UBC#807; UBC#811; A18; UBC#840; A07; UBC#889; UBC#835; ERF-F; MYB; 1962; B01; A19;
- (3) 10- D_L – UBC#807; UBC#811; A18; UBC#840; A07; UBC#889; UBC#835; ERF-F; MYB; 1962;
- (4) 5- D_L – UBC#807; UBC#811; A18; UBC#840; A07;
- (5) 3- D_L – UBC#807; UBC#811; A18;
- (6) 1- D_L – UBC#807.

For each of the primer combinations, the main indices of genetic diversity (P_B , H_e , S and D_j) were calculated, and analysis of molecular variance (AMOVA) was performed (Table 2). The H_e ranged from 0.143 to 0.227 and the S varied between 0.216 and 0.341, when different primer combinations were used. The use of primers with the highest D_L values in all cases resulted in increased genetic diversity indices compared with all primers, apparently due to the growth of the portion of highly polymorphic primers. The values of the indices also slightly increased with the further reduction of number of primers used in analysis,

Table 1. Primers with the highest values of discriminating power (D_L) selected for the use in population genetic studies of *Gentiana lutea*

	Primer	R_p^a	D_L^a	ND ₃₀	N_t^b	ND ₈₆ estimated	ND ₈₆ experimental
1	UBC#807	11.9	0.967	0	32	0	4
2	UBC#811	12.3	0.964	1	38	8.4	3
3	A18	7.7	0.956	5	21	42.0	52
4	UBC#840	7.2	0.953	5	23	50.4	21
5	A07	5.4	0.953	6	26	50.4	24
6	UBC#889	8.7	0.953	7	26	50.4	20
7	UBC#835	4.7	0.951	7	25	58.8	16
8	ERF-F	9.1	0.949	8	23	67.2	56
9	MYB	7.7	0.949	8	22	67.2	106
10	1962	6.1	0.949	8	17	67.2	58
11	B01	5.5	0.944	10	20	84.0	34
12	A19	7.3	0.942	11	18	70.2	28

R_p , resolving power; ND₃₀, experimental value number of non-differentiated pairs determined for the sample of 30 plants; N_t , total number of generated bands; ND₈₆, ND for the sample of 86 plants estimated from the frequencies of banding patterns generated in the sample of 30 plants (ND₈₆ estimated) or determined experimentally (ND₈₆ experimental).

^aValue calculated for the sample of plants of two populations ($n = 30$). ^bValue calculated for the sample of plants of six populations ($n = 86$).

Table 2. Measures of genetic diversity in two populations of *Gentiana lutea* (average data) obtained using different combinations of primers with the highest values of D_L

Primers combination	P_B (%)	H_e	S	D_j average (%)	AMOVA (within/between populations) (%)
40 Primers	44.9	0.143 ± 0.006	0.216 ± 0.009	42.1	37/63
12- D_L	56.8	0.178 ± 0.010	0.270 ± 0.014	53.0	48/52
10- D_L	56.7	0.179 ± 0.010	0.272 ± 0.015	52.1	48/52
6- D_L	57.0	0.174 ± 0.013	0.267 ± 0.019	58.7	47/53
3- D_L	60.8	0.183 ± 0.017	0.281 ± 0.024	67.5	48/52
1- D_L	66.7	0.227 ± 0.033	0.341 ± 0.047	66.8	49/51

P_B , proportion of polymorphic bands; H_e , expected heterozygosity; S , Shannon index; D_j , Jaccard's genetic distances; AMOVA, analysis of molecular variance.

with the exception of small decrease in H_e and S observed when using combination of six primers. However, this variation was insignificant when using different primer sets, except for the combination 1- D_L .

AMOVA revealed significant genetic divergence of the populations, the difference between populations accounted for no less than 50% of total genetic variation. Decrease in the number of primers from 40 to 12 or less led to the change in distribution of genetic diversity towards the growth of variation within population. This is obviously associated with a specific distribution of polymorphic bands revealing mainly the differences among populations, which seem to be generated mostly by the primers with small percentage of polymorphic fragments. Nevertheless, the results of AMOVA did not vary significantly when using different number of selected primers (Table 2). Therefore, the use of any number of primers from 3 to 12 for assessment of genetic diversity is expected to give the comparable results.

To further test the informativeness of selected primers, as well as the effectiveness of discriminating power (D_L) as an index of primer informativeness, we used them to analyse a set of 86 *G. lutea* plants from six populations from the Sydovets and the Chornohora ridges of the Ukrainian Carpathians. For each primer, we compared the experimental number of non-differentiated pairs with the value estimated from the frequencies of banding patterns generated in the set of 30 plants (see ND₈₆ in Table 1). The estimated value was greater than the experimental one for most of the primers with the exception of UBC#807, A18 and MYB. An increase in the total number of fragments and proportion of polymorphic fragments resulted from the inclusion of additional plants of other populations in the study may be one of the reasons for this. We found strong correlation ($r_s = 0.75$, $P = 0.05$) between the number of non-differentiated pairs of plants estimated from the analysis of polymorphism in the sample of plants from two populations ($n = 30$) and experimental value obtained for

the sample of plants from six populations ($n = 86$). These results support the effectiveness of discriminating power as an index for evaluation of primers as well as demonstrate high informativeness of selected primers.

To determine the minimum number of primers (markers) required to resolve genetic relationships among *G. lutea* accessions from six populations, we constructed the UPGMA dendrogram based on the matrix of D_j and estimated bootstrap support for key nodes, using for the analysis the data from different number of markers (Fig. 2(a)). Pairwise genetic distances between individual plants were calculated from the data generated using selected primers with the highest D_L values. The analysis started with the primer UBC#807, and then the number of markers was gradually increased by adding data from successive primers listed in Table 1. Samples from the individual populations were grouped in the dendrogram together into clearly separated clusters. There was also no obvious grouping of populations from the same mountain ridge even in the case of relatively small distance between them.

Bootstrap support was calculated for six key nodes that represent the clusters of individual populations. Relationship between the number of markers used in the analysis and the average bootstrap support across the key nodes is shown in Fig. 2(b). The average bootstrap value increased as the data were added to the analysis, and bootstrap support for all nodes of interest approached 90% even when the first three primers producing 91 markers in total were used, while six primers providing 166 markers gave the bootstrap support above 99%. Thus, only three selected primers with the highest D_L generate the number of polymorphic PCR markers sufficient to assign the *G. lutea* plants to their population of origin with accuracy of more than 85%. Calculations made from the data of discriminating power show that combination of any three primers with the highest D_L , under assumption of independence of the primers pattern, is theoretically enough to discriminate between all genotypes in a set of about 1000 *G. lutea* plants (Table 3).

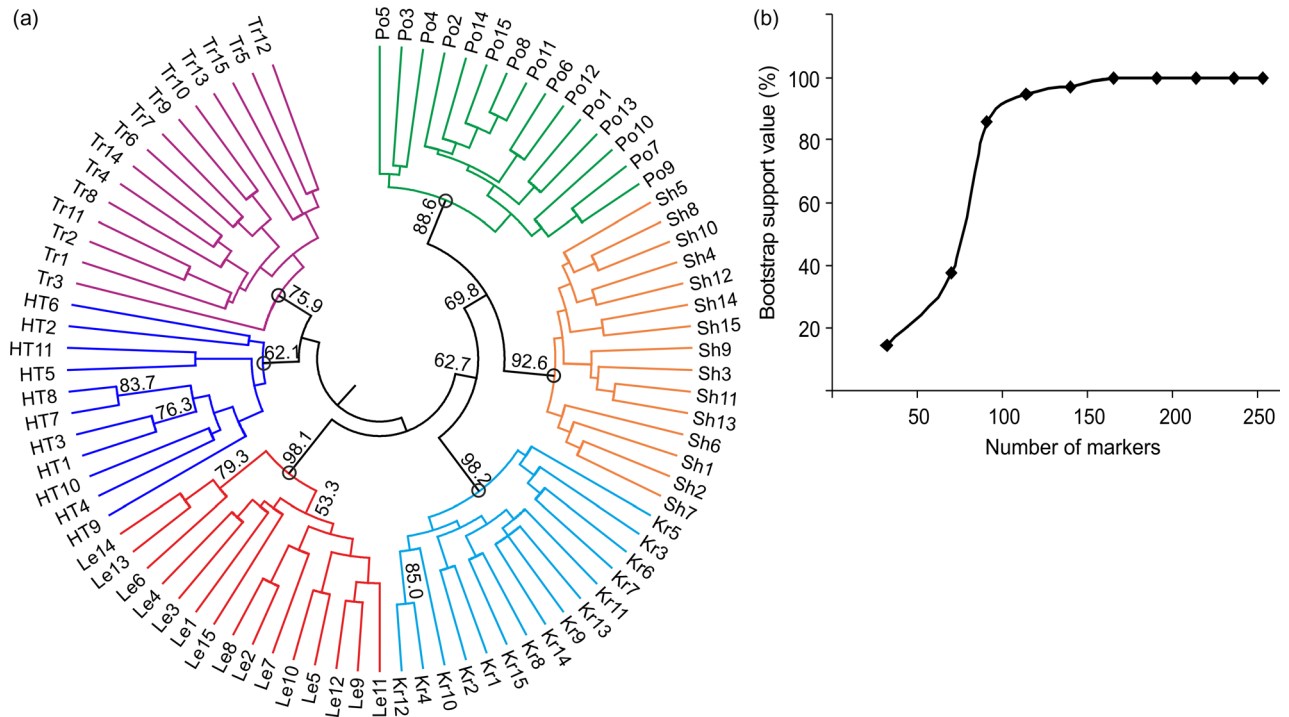


Fig. 2. (a) Relationships among 86 *Gentiana lutea* genotypes of six populations estimated from the data of 91 markers generated by three most informative primers (UBC#807; UBC#811; A18). The tree was constructed using the unweighted pair group method with arithmetic mean (UPGMA) algorithm based on Jaccard's genetic distances between individuals. Bootstrap values, calculated based on 1000 replicates, are presented for the six nodes of interest (circled). Bootstrap values greater than 50% are also shown. (b) Relationship between the number of markers used in analysis and the average bootstrap support across six key nodes including individual populations on UPGMA dendrogram. For the description of abbreviations, refer Fig. 1 legend.

Discussion

As a result of the study, we have selected the effective primers for the use in population genetic analysis of a rare plant species *G. lutea* and demonstrated their utility for the assessment of genetic diversity and analysis of genetic structure of the species populations. Several types of PCR-based markers were chosen for the study, including those that can be directly associated with protein coding regions of functionally important genes (such as CDDP and RGAP primers that are targeting conserved sequences of disease resistance genes and stress response genes). Our choice of the primers was based on consideration of possible use of the PCR-based marker system developed for ecogenetical investigations and studies on the adaptive genetic variation associated with environmental conditions.

Two indices were used to evaluate informativeness of individual primers, resolving power (Prevost and Wilkinson, 1999) and discriminating power (Tessier *et al.*, 1999). Both of them take into account the distribution of polymorphic alleles in studied set of genotypes. While R_p is related to the distribution of individual amplified fragments, D_L is calculated from the frequencies of banding patterns generated by a primer. Comparative

analysis of the data revealed significant yet only moderate relationship between these indices as well as between R_p and the number of non-differentiated pairs. This is primarily due to the fact that the R_p calculation does not take into account potential non-independent distribution of bands in profiles of individual plants when the generation of one of the fragments is strongly associated with the presence of the other. The discriminating power is free from this limitation and directly associated with the number of pairs non-differentiated by a primer. The obtained results indicate that the resolving power is

Table 3. Theoretical efficiency of various primer combinations calculated under hypothesis of independence of their patterns for 500 and 1000 accessions

Primer combination	ND ₅₀₀ estimated	ND ₁₀₀₀ estimated
UBC#811 + A18	3.3	13.2
UBC#811 + UBC#889	3.9	15.8
UBC#811 + A18 + A07	0.1	0.2
UBC#811 + A18 + UBC#889	0.1	0.2
UBC#811 + A18 + UBC#840	0.1	0.2
UBC#811 + A18 + UBC#835	0.1	0.2
UBC#811 + UBC#840 + A07	0.1	0.2

not always a reliable index of primer informativeness, namely the ability to discriminate between the different genotypes. This is especially true in the case of multilocus markers. The same conclusion was done by Saini *et al.* (2010) as a result of comparison of several different measures of PCR primers efficiency. They found that D_L was the most effective index of primer informativeness in selection of molecular markers for identification of mung bean varieties. This index was also used successfully to develop algorithm and computer software for selection of minimal sets of molecular markers for accessions and variety identification (Caroli *et al.*, 2011; Fujii *et al.*, 2013). Moreover, Fujii *et al.* (2013) showed that the use of discrimination power in calculations may significantly accelerate computation speed when a large number of markers and varieties are involved.

Informativeness of the primers that were selected based on D_L value was further confirmed using a larger set of plants from six populations located on two mountain ridges separated by a river valley of the Ukrainian Carpathians. Comparison of the number of non-differentiated pairs experimentally estimated and theoretically calculated from the frequencies of banding patterns revealed remarkable agreement between these values. Moreover, for some primers, the real number of non-distinguishable pairs turned out to be less than expected. These results indicate that discriminating power may be used to evaluate the informativeness of primers intended for the use in assessment of genetic diversity of the species. It allows selecting primers, which can be used both to identify individual genotypes and to differentiate between individual populations.

In total, 12 primers with the highest values of D_L were selected for the use in further studies. However, calculations made from the data of discriminating power show that as little as three of these primers are enough to differentiate individual genotypes in a group including as many as 1000 *G. lutea* plants. Moreover, we have demonstrated that the use of different number of selected primers from 3 to 12 gives comparable measures of genetic diversity. Nevertheless, it should be mentioned that the values of such indices as the H_e , S and genetic distance between individuals depends largely on the informativeness of used markers. In particular, when the most informative primers were used, these indices were 21–58% higher than when all the 40 screened primers were included in the analysis. This emphasizes the need to consider measures of informativeness of markers applied in the analysis when comparing estimates of genetic diversity made in different studies.

The analysis of relationship between the number of markers used in the study and the average bootstrap support for key nodes on dendrogram of 86 plants from six populations showed that the number of bands

produced by three of the primers with the highest D_L was sufficient to give bootstrap values over 85%, but with the use of six primers the average bootstrap value exceeded 99%. Obviously, an increase in the number of analysed genotypes will also require an increase in the number of primers used in the analysis. The optimal number of markers for use in molecular genetic analysis depends also on the goal of the study. For assessment of genetic diversity, surveys of population structure, genetic relatedness or assignment studies, there is usually an optimal number of markers that provide adequate statistical significance for obtained results and further increase in the number of markers does not necessarily improve the results of analysis. On the other hand, a search for loci exposed to natural selection or associated with a specific trait requires the use of as many markers as possible. Thus, a minimal set of three to six selected primers can be sufficient for quick assessment and subsequent monitoring of genetic diversity of *G. lutea* populations, depending on the sample size and degree of differentiation between populations, while the rest of the primers with the D_L values above 0.8 may be used for ecogenetic surveys.

On the whole, the results of our preliminary studies carried out with selected primers demonstrated moderate level of genetic diversity within the samples from two populations as well as clear differentiation of all studied populations. This is evidenced by the data of AMOVA according to which among-population variation accounts for over 50% of total variance and grouping of the samples from individual populations in UPGMA dendrogram into the distinct clusters. In the future, we plan to conduct a more detailed analysis of genetic variation in the populations of *G. lutea* considering population and environmental parameters.

Supplementary material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S147926211400104X>

Acknowledgements

We thank Iryna Petrusha (Foreign Languages Department, Institute of International Relations, Kyiv National Taras Shevchenko University) and Volodymyr Adonin (Institute of Molecular Biology and Genetics of NAS of Ukraine) for assistance in translating the text into English. This study was financially supported by the National Academy of Sciences of Ukraine through the Targeted interdisciplinary programme of scientific research ‘Fundamentals of molecular and cellular biotechnology’ and from the

Ministry of Education and Science of Ukraine in a framework of the project 'Physiological, ecological and biotechnological bases of some *Gentiana* L. species conservation *in vitro* and *in situ*'.

References

- Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314–331.
- Caroli S, Santoni S and Ronfort J (2011) AMAcAID: a useful tool for accurate marker choice for accession identification and discrimination. *Molecular Ecology Resources* 11: 733–738.
- Collard BCY and Mackill DJ (2009) Conserved DNA-derived polymorphism (CDDP): a simple and novel method for generating DNA markers in plants. *Plant Molecular Biology Reporter* 27: 558–562.
- Dong P, Wei Y-M, Chen G-Y, Li W, Nevo E and Zheng Y-L (2009) Resistance gene analog polymorphisms (RGAPs) in wild emmer wheat (*Triticum dicoccoides*) and their ecological associations. *Genetic Resources and Crop Evolution* 56: 121–136.
- Fujii H, Ogata T, Shimada T, Endo T, Iketani H, Shimizu T, Yamamoto T and Omura M (2013) Minimal marker: an algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient variety identification. *Journal of Bioinformatics and Computational Biology* 11: 1250022. doi: 10.1142/S0219720012500229.
- Jensen SR and Schripsema J (2002) Chemotaxonomy and pharmacology of Gentianaceae. In: Struwe L and Albert VA (eds) *Gentianaceae: Systematics and Natural History*. Cambridge: Cambridge University Press, pp. 573–631.
- Konvalyuk II, Mel'nyk VM, Drobyk NM, Kravets NB, Twardovska MO and Kunakh VA (2011) RAPD- and ISSR-analysis of genetic variation in *Gentiana pneumonanthe* L. tissue and organ culture. *Visnik Ukrain'skogo Tovaristva Genetikiv i Selekcioneriv* 9: 2–31.
- Mayorova OY, Grytsak LR, Terehova GI, Mel'nyk VM, Andreev IO and Drobyk NM (2013) *Gentiana lutea* L. (Gentianaceae) in the flora of Ukrainian Carpathians: characteristics and current status of populations. *Ukrains'kyj Botanicnyj Zurnal* 70: 780–787.
- Mel'nyk VM, Spiridonova KV, Andreev IO, Strashniuk NM and Kunakh VA (2004) Variability of nuclear 18S-25S rDNA of *Gentiana lutea* L. in nature and in *in vitro* tissue culture. *Cytology and Genetics* 38: 14–19.
- Mosula MZ, Konvalyuk II, Mel'nyk VM, Andreev IO, Bublyk OM, Drobyk NM and Kunakh VA (2013) Genetic diversity in *Gentiana lutea* L. populations from Svydivets ridge of the Ukrainian Carpathians. *Visnik Ukrain'skogo Tovaristva Genetikiv i Selekcioneriv* 11: 250–259.
- Mosula MZ, Konvalyuk II, Mel'nyk VM, Bublyk OM, Andreev IO, Drobyk NM and Kunakh VA (2014) Analysis of genetic diversity in *Gentiana lutea* L. populations by method of marking between retrotransposon's sequences (IRAP-PCR). *Fiziologija Rastenii i Genetika* 46: 45–55.
- Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.
- Peakall R and Smouse PE (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537–2539.
- Powell W, Morgante M, Andre C, Hanfey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225–228.
- Prevost A and Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theoretical and Applied Genetics* 98: 107–112.
- Rogers SO and Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology* 5: 69–76.
- Saini M, Singh S, Hussain Z and Sikka VK (2010) RAPD analysis in mungbean [*Vigna radiata* (L.) Wilczek.] II: a comparison of efficiency parameters of RAPD primers. *Indian Journal of Biotechnology* 9: 276–282.
- Schluter PM and Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. *Molecular Ecology Notes* 6: 569–572.
- Strashniuk NM, Les'kova OM, Zahrychuk HYa, Mel'nyk VM and Kunakh VA (2006) Biologically active substances of *Gentiana* L. genus. 1. Biosynthesis and physiological influence. *Phytotherapy* 1: 31–41.
- Tessier C, David J, This P, Boursiquot JM and Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theoretical and Applied Genetics* 98: 171–177.
- Twardovska MO, Strashniuk NM, Mel'nyk VM, Konvalyuk II and Kunakh VA (2009) RAPD-analysis of the genome polymorphism for some *Gentiana* L. species from the Ukrainian flora. *Reports of the National Academy of Sciences of Ukraine* 5: 200–204.
- Yap IV and Nelson RJ (1996) WinBoot: a program for performing bootstrap analysis of binary data to determine the confidence of UPGMA-based dendrograms. *IRRI Discussion Paper Series*, No. 14. Manila, Philippines: International Rice Research Institute.